Non-Religious Ethics? A critical notice of Derek Parfit, On What Matters

Non-Religious Ethics?

A critical notice of Derek Parfit, *On What Matters*

The History Workshop was a movement of radical social historians which flourished in Britain in the 1960s and 70s. Its goal was to promote "history from below" – to tell the stories of those left out of conventional narratives, and, at the same time, to open up the practice of history itself. Anyway, when the group decided to start a journal, there was a debate over whether it should carry book reviews. Weren't book reviews – the ranking of others' work, delivered in a tone of apparent omniscience – examples of exactly the kind of academic gate-keeping against which they had set themselves? So, in its early issues at least, the *History Workshop Journal* didn't carry reviews but "Enthusiasms" – contributions in which someone would introduce a new book to readers in positive terms, without pretending to be marking it in some transcendental Prize Fellowship Examination.

I have always had a lot of sympathy with their point of view and it came back to me with force when I was asked to provide a critical notice of *On What Matters*.[1] *On What Matters* is a vast book (it's been said that it is the only work of contemporary moral philosophy that is visible from space) and it challenges its readers in many ways, one of which – admittedly not, perhaps, the most significant – is that it is impossible to review conventionally. Should the reviewer summarize the book? Well, it deals with most of the central questions of normative and meta-ethics. It also offers an extended interpretation of Kant's moral philosophy and addresses some of the major questions of epistemology and metaphysics for good measure. Even a long

[1] *On What Matters*, 2 volumes (Oxford: Oxford U.P., 2012). References to *On What Matters* will be abbreviated as *OWM*, followed by volume and page number.

critical notice would be hard pressed to go past such simple headlines and slogans if it aimed to give a comprehensive survey.

As for assessment, that seems presumptuous. That *On What Matters* is the major work of a highly respected and influential contemporary author is, surely, obvious. Will it pass "the test of time"? My guess is that its size and scope will make *On What Matters* an unavoidable part of the landscape of moral philosophy for the foreseeable future. How it will be seen, however, depends, at least in part, on what direction the discipline itself takes in future years. Although Parfit is exercising a voice – a loud one – in trying to influence which way moral philosophy should go, how far it will follow his prescriptions is, surely, impossible now to predict. Will the moral philosophers of the future find *On What Matters* a towering peak on which to stand and survey the territory of their subject or a boggy slough through which they must struggle before reaching firmer ground? To be quite honest, I have no clue.

Yet the idea that one should just recommend a work and leave the rest to the reader seems inadequate – at least, in the case of philosophy. Works of philosophy are arguments, attempts to persuade their audience. Surely that requires more than a wave of recognition – or even a cheer of welcome. And doesn't the reader need some help in approaching texts as large and ambitious as *On What Matters*? A work of this magnitude deserves engagement from fellow-philosophers – public engagement that will, ideally, encourage others to continue the task.

I do not intend to shirk that duty, but, before I embark on it, I want to make a methodological point in preparation. From its outset, Western philosophy has been

haunted by the example of mathematical proof as an ideal of rigorous reasoning, something that in turn encourages the assumption that evaluating a philosopher's argument should be a matter of checking through a chain of reasoning to see where the links are weak. But the differences are at least as important. Even granting that philosophical arguments are deductive in form, we need to assess the truth of their premises. Moreover, philosophical arguments are characteristically enthymematic – that is, they need further, unstated premises in order to be conclusive. So engaging with a work of philosophy involves identifying and evaluating many things the author does not express directly.

Not that Parfit is unaware of this. The attempt to make explicit and defend his assumptions draws him into addressing many areas of the discipline. *On What Matters* relates discussions of substantive ethical questions to meta-ethics and moves thence to epistemology, metaphysics and the philosophy of mind. But those issues in turn point still further: in the end, the whole of philosophy is a single extended, interconnected web. It is, I think, the recognition of this holism that gives the book's extreme length its best justification. Yet, even were it twice as long, it couldn't examine all of the assumptions that are, at least implicitly, connected with its central themes. Moreover, whether such assumptions are questionable or not is itself a matter of legitimate philosophical disagreement.

For that reason, any critical engagement will inevitably be partial and (in the sense that it reflects the writer's own understanding of the subject) personal. It will become apparent that many of my assumptions are very different from Parfit's. Although (of course) they are ones that seem compelling to me, I do not want to pretend that they

are beyond reasonable disagreement – hence my distrust of the reviewer's conventional tone of superiority. On the contrary, the following reflections are offered in a respectful spirit of dialogue in the hope that, even if they are not persuasive, the readers of this journal (and, if possible, the author himself) will find them thought-provoking.

II

I want to start, not at the beginning of *On What Matters* but, in a sense, immediately before it, at the end of *Reasons and Persons*, Parfit's previous book. The passage that I have in mind (it is, in fact, the book's concluding paragraph) is, to me, extremely striking.

> Belief in God, or in many gods, prevented the free development of moral reasoning. Disbelief in God, openly admitted by a majority, is a recent event, not yet completed. Because this event is so recent, Non-Religious Ethics is at a very early stage. We cannot yet predict whether, as in Mathematics, we will all reach agreement. Since we cannot know how Ethics will develop, it is not irrational to have high hopes.[2]

I find this passage so provocative and radical that exploring it might easily occupy the rest of my space. The picture it gives is the exact opposite of the one that is most familiar nowadays. It is commonly assumed that, as belief in God has receded –

[2] Derek Parfit, *Reasons and Persons* (Oxford: Oxford U.P., 1984), p. 454

become tentative and problematic, at least, even if it has not actually disappeared – it has left us in an age of uneasy pluralism. Not only do people now disagree about moral matters, they disagree about that disagreement. Notoriously, some people regard as "immoral" behaviour that others see as just a matter of taste or preference. Worse, there seems to be no common framework within which even disputes that are agreed to be moral can be adjudicated.

A picture of this kind is memorably evoked at the beginning of Alasdair MacIntyre's *After Virtue*, but the sense of disintegration in contemporary moral life is not confined to those who lament the passing of religiously-inspired ethics. Thus Simon Blackburn, whose highly critical review of *After Virtue* is written from a standpoint as far removed from MacIntyre's as possible, talks of "our sense of the fractured, fragmentary nature of moral requirements" (he thinks, contra MacIntyre, however, that this is nothing new).[3] Contrast this with the confidence that people have in the natural sciences, even in areas – the operation of the human brain is a striking example – in which our current knowledge is pitifully small. It is, at the least, unusual to discover someone who thinks that morality might be set on that same path.

It is true, Parfit might concede, that people nowadays disagree passionately about some moral matters. But perhaps we shouldn't be distracted by that fact; we should look instead at the very many things that we agree about – things that don't get us upset but which we simply take for granted. Thus, one might point out, no one now seriously argues that slavery is permissible. The fact that slavery is wrong is as much part of our agreed beliefs as the fact that the earth goes round the sun. Yet that belief

[3] *Philosophical Investigations*, *5* (2) 1982: 146-53, p.152

was once far from common. The ancient world, in fact, had an equally pervasive opposite conviction: that slavery was a justifiable feature of human society, however regrettable. Of course, it seems to me beyond question that the change from that shared judgement to our own is a huge moral improvement, but it is hard to believe that the mere agreement in judgement itself is greater now than it was in times past.

Is it just the continued presence of religion that lies behind moral disagreement? There is no doubt that the most vehement moral conflicts in contemporary Western societies are ones in which religion plays a part – questions of bioethics come to mind. Yet it seems over-simple to think that it is only the religious and the non-religious (and the religious amongst themselves) who disagree on such issues. Of course, at present, the number of explicitly non-religious participants in public moral debates is limited, but do even they agree? Problems of life and death divide the secular, as do issues of justice and equality.

The reason that Parfit presents for the assertion that it is religion that gets in the way of moral agreement is also questionable. It is only natural, he says, for religious believers to think that we ought to obey God's commands (*OWM*, 2, 553). But that is to assume that religious ethics is essentially a matter of obeying commands whose authority consists in the fact that they have come from God (and that there are conflicting understandings of what those commands amount to). Yet, on the dominant self-understanding of Western Christianity, God is not an arbitrary source of imperatives. On the contrary, God's commands coincide with what it is open to human beings independently to recognize as good. So, from the point of view of orthodox religion, *pace* Parfit, there is no reason why belief in God should have

prevented “the free development of moral reasoning”. On the contrary, religious ethics, no less than secular, takes itself to draw on the inherent powers of human reason (that is why Catholic ethics felt free to take so much of its content from Aristotle). If that is so, then the idea becomes much less plausible that we have recently moved from “Religious” to “Non-Religious Ethics” and that the path to agreement, once blocked by religious belief, is now clear.

Parfit endorses Sidgwick’s belief that “ethical science” should be pursued “by an application to it of the same disinterested curiosity to which we chiefly owe the great discoveries of physics” (*OWM*, 2, 153). This suggests another possible explanation for the apparent fact of widespread moral disagreement. There is often a strong divergence between beliefs formed on the basis of scientific evidence and reflection and ones held in the population at large. Many citizens of the world’s most powerful economy (and greatest producer of greenhouse gas emissions) do not believe in the existence of global warming. Moreover, politicians across the Western world currently propagate the notion that cutting public spending is an appropriate policy in economies suffering from high unemployment and slow growth. Yet the popular credence given to such beliefs should not shake the confidence of those who study those issues objectively that they are quite wrong. Perhaps the existence of popular moral disagreement could be compatible with progress at the level of “ethical science” in the same way.

That would require us to reflect more deeply on what the method of “ethical science” might be. In climate science or economics, we have appropriate ways of testing theories empirically, ones that give us confidence in the beliefs based on them even

when those beliefs are rejected by many people. I shall return to this issue below, but, for the present, let me note the following. The method of ethics is grounded, Parfit, following Sidgwick, asserts, in intuition: "We have *intuitive* abilities to respond to reasons and to recognize some normative truths" (*OWM*, 2, 544). If that is so, however, then it is puzzling how there could be the same kind of discrepancy between "popular" and "scientific" perspectives that one might find in (say) physics. Is it that some people's intuitions are hopelessly clouded and others' not? More importantly, from what vantage point could we tell who is who?

Nevertheless, there is moral truth, Parfit believes, and it is the business of philosophers to discover it (and to defend its status *as* truth against subjectivists of various kinds). For this to be so, that truth must be unique, even if it is structurally complicated. As he writes in response to Susan Wolf:

> … it *would* be a tragedy if there was no *single true morality*. And conflicting moralities could not all be true. In trying to combine these different kinds of moral theory, my main aim was not to find a supreme principle, but to find out whether we can resolves some deep disagreements. As Wolf claims, it would not matter greatly if morality *turned out* to be less unified, because there are several true principles which cannot be subsumed under any single higher principle. But if we cannot resolve our disagreements, that would give us reasons to doubt that there are *any* true principles. There might be nothing that morality *turns out to be*, since morality might be an illusion. (*OWM*, 2, 155)

Hence the importance of convergence.

On What Matters is based on the Tanner Lectures given by Parfit at Berkeley under the title *Climbing the Mountain*. Parfit does not explain why he changed the title (perhaps he simply got tired of all the feeble jokes it inspired) but he still retains the thought behind the metaphor: you may start from different positions, but the point you will reach if you ascend far enough is the same. The central part of the book is an attempt to show that three apparently quite different approaches to moral philosophy – Kantianism, contractualism and consequentialism – come together in this way and it is to this that I now turn.

III

Parfit's reading of Kant takes up a great deal of this very long book. To respond to it in the space available I must put things very schematically and ask the reader to believe that there is supporting evidence for the claims I make, even if it is not possible to present that evidence fully here. Parfit's fundamental interpretive claim is extremely radical but, fortunately, also extremely concisely formulated: it is that "Kantian Contractualism implies Rule Consequentialism" (*OWM*, 2, 153). Understanding and evaluating this claim is, however, more complicated. For a start, Parfit's Kant is a deeply inconsistent thinker.

"When I became obsessed with Kant," he writes in the Preface, "I tried to restate more clearly some of Kant's main claims and arguments, and found this task very frustrating. I couldn't fit Kant's claims together in a coherent view, and many of

Kant's arguments seemed to be obviously invalid or unsound…. It would have helped me to know that Kant did not have a single coherent theory. When we ask whether Kant accepts or rejects some claim, the answer is often 'Both'." (*OWM*, 1, xli-xlii) According to Parfit, Kant's inconsistency and his originality are essentially connected. "The truth is that in the cascading fireworks of a mere forty pages, Kant gives us more new and fruitful ideas than all the philosophers of several centuries. Of the qualities that enable Kant to achieve so much, one is inconsistency." (*OWM*, 1, 183)

Yet, if this is so, what does it mean to say that Kant's ideas *imply* ones that are usually thought to be very different? A logician will tell us that inconsistency – a set of sentences that cannot all be true together – implies *anything*; it could just as easily be perfectionism or act consequentialism as rule consequentialism. Parfit's claim is clearly about argumentative structure, however: that, if we take *what is fundamental* to Kant as our starting point and follow it through *where it ought to go*, then it will lead us to Rule Consequentialism. But in that case the interpretive situation we face is much more problematic, for we must decide what is central and what is peripheral in Kant. How is that to be done?

Parfit dismisses very many things that Kant himself obviously thought to be essential to his project. He has no sympathy with Kant's theory of freedom and the deep connection that Kant proclaims between the metaphysics of transcendental idealism and morality. He rejects most of Kant's specific moral judgements (for example, those about lying, suicide and punishment – he does not, to my recollection, discuss Kant's views on sexuality, though I can hardly imagine that he would find them any the less repellent). Above all, he is unqualifiedly hostile to Kant's ideas about desert and

happiness – the idea that happiness is only good if deserved and that, correspondingly, deserved suffering is good.

So what then does Parfit himself take to be fundamental? The answer is apparent even in his very brief statement of his central claim: it is "Kantian *Contractualism*", as he calls it, that implies Rule Consequentialism. Parfit introduces his main discussion of Kant in the following words:

> According to Kant's best-loved principle, often called *The Formula of Humanity*: We must treat all rational beings, or persons, never merely as a means, but always as ends. To treat people as ends, Kant claims, we must never treat them in ways to which they could not consent. (*OWM*, 1, 177)

Yet the versions of consent theory (like Parfit, I shall use "contractualism" and "consent theory" interchangeably) that Parfit finds in Kant do not satisfy him. After a great deal of detailed argument, the Kantian formula is revised to become "Everyone ought to follow the principles that everyone could rationally will to be universal laws" (*OWM*, 1, 407). Even after all of this argument, however, the actual content of the formula still remains to be settled. Which principles can be rationally willed by everyone? It is here that consequentialism enters. Only consequentialism gives us a principle that everyone has a sufficient reason to choose, he claims. So, in the end: "Everyone ought to follow the principles whose being universal laws would make things go best, because these are the only principles whose being universal laws everyone could rationally will." (*OWM*, 1, 25) That is the point – the "Triple Theory" as he calls it – at the peak of the mountain.

Well, OK, *c'est magnifique mais* … is it Kant? If Parfit is right, then pointing to elements of Kant's theory that are discrepant with his interpretation will not be dispositive. Kant, after all, on Parfit's reading, is wildly inconsistent. So let me ask instead whether the starting point itself is plausible. Is Kant really a "contractualist"?

To return to the chapter in which Parfit opens his account of Kant, the passage continues:

> In explaining the wrongness of a lying promise, for example, Kant writes "he whom I want to use for my own purposes with such a promise cannot possibly agree to my way of treating him".

The quotation comes from the *Groundwork* (*Ak.* 4, 430)[4] and it certainly sounds as if, here at least, Kant is employing some form of consent theory. Yet even this does not persuade me.

The German word that is being translated as "agree to" is "*einstimmen"*. "*Einstimmen*" is indeed properly translated as "agree". But agreement between people can take more than one form. You and I can agree with one another in one way by arriving at the same judgement – just as our watches can agree with one another. By

[4] References to Kant are to the Collected Works published by the Prussian Academy of Sciences (*Akademie-Ausgabe*) and are referred to as *Ak.* followed by volume and page numbers. Widely-used English translations contain Academy Edition page numbers in the margin. The exception is the *Critique of Pure Reason*, for which the convention of reference to the page numbers of the first and second editions, preceded by 'A' or 'B' respectively, is followed.

contrast, I can agree with you in another way by *agreeing to* something that is proposed. The former, we might say, is the agreement of coincidence, while the latter is the agreement of consent. Now "*Einstimmung*" means the agreement of coincidence, not the agreement of consent. That is how Kant consistently uses it throughout his writings. For instance, he describes judgements as standing in "*Einstimmung*" or "*Widerstreit*" – agreement or conflict – in the *Critique of Pure Reason* (B320). The sentence in question should properly have been translated as "*agree with*" not "*agree to*" my way of treating him. There is, in fact, a different German word which appropriately translates "consent" – *zustimmen*, which means directly "assent to". *Einstimmen* is used very rarely in the moral writings (Parfit's quotation is its sole appearance in the *Groundwork*) and *zustimmen* not at all. In short, the direct textual evidence for Kant being a contractualist is either slim (if you agree with the Cambridge translation of *Ak.* 4, 430) or non-existent (if you take my point).

But in taking Kant to be a consent theorist, Parfit is following some of the most respected of Kant's modern advocates. Thus, after the quotation from *Ak.* 4, 430, Parfit's text continues with comments on that passage from two leading Kant interpreters, Christine Korsgaard and Onora O'Neill, who discuss it in terms of "the conditions of possible assent" and of the requirements of "genuine consent" respectively. If Kant is not explicitly a consent theorist, why is the belief that he is one implicitly so pervasive amongst (at least, contemporary English-language) writers on Kant?

Apart from the *Formula of Humanity*, the *Groundwork* also contains the *Formula of Universal Law*, which states that one should "act only in accordance with that maxim through which you can at the same time will that it become a universal law" (*Ak*. 4, 421). Perhaps it is here that Kant establishes himself as a consent theorist. Obviously, if we will something as a universal law then it is binding on everyone who is a moral agent and, for that reason, is something that every such agent (a "lawgiver in the kingdom of ends") should will. Yet any acceptable moral principle ("Obey the Ten Commandments", say) will have that quality. "Willing a maxim as a universal law" just seems to be a way of stating in Kantian language that we assert the principle ("maxim") in question. The issue is how whatever it may be that we can (or can't) will plays a role in settling the substance or content of that law. If we look at what Kant says about this issue in the *Groundwork,* however, it does not look as if the idea of consent plays any role. In the case of strict duties, he writes, "their maxim cannot even be *thought* without contradiction as a universal law of nature" while, in the case of wide duties, it is "impossible to *will* that their maxim be raised to the universality of a law of nature because such a will would contradict itself" (*Ak.* 4,424). So it is either a matter of a "contradiction in thought" or of a "contradiction in will". But, if the question is: can we consent to some possible rule?, then it must, surely, be, at least, *thinkable*. On the other hand, the examples that Kant gives for actions that embody a "contradiction in will" are suicide and the failure to develop one's talents – implausible candidates, one would think, for things to which one could not consent. I wouldn't say for a second, of course, that it is easy to find a way to interpret either of those two ideas – simply that there is nothing in the text itself to support the belief that they must be interpreted in contractarian terms.

There is, however, one area of Kant's ethical thought that does seem to employ something like a consent theory: his writings on politics, above all in the *Metaphysics of Morals*. Take what Kant calls the "universal law of right" (*das allgemeine Rechtsgesetz*) and which he formulates as follows: "so act externally that the free use of your will can exist together with the freedom of everyone in accordance with a universal law" (*Ak*. 6, 231). The "will" that Kant refers to here is the *Willkür*, sometimes translated (not very felicitously) as the "elective will", but meaning, at any rate, that power of choice that can be exercised arbitrarily, in any direction. For the *Willkür* of everyone to co-exist there must be a framework that accommodates people's different choices, whatever they might be.

Yet I think that the presence of such themes in Kant's writings on politics reinforces my point. It reminds us of the dualism of Kant's picture of moral life – the contrast between the external realm of law and politics on the one hand and the inner world of virtue on the other. Morality, of course, has to include both, but any principle that depends on the idea of a hypothetical contract between parties who choose on the basis of giving weight to their self-interested claims is bound, I fear, to skew us towards the former, at the expense of failing to account for the latter. *Wille* (the positively free, moral will) becomes assimilated to *Willkür* and a principle that would make sense in relation to the latter (for, example, self-interested choice behind a veil of ignorance) is treated as if it is what underlies the project as a whole. In consequence, ideas such as "duties to oneself" (which Kant says explicitly "take first place, and are the most important duties of all" (*Ak.* 27, 341)) are treated as peripheral

to what is asserted to be Kant's fundamental enterprise, while, at the same time, Kant is said to lack a coherent overall theory.

IV

In arguing that Kant is not a contractarian I by no means wish to deny that values based on choice have an important place in Kant's thought – of course they do. It is just that they do not have the foundational role in Kant's moral system that a contractarian approach assigns to them. But, if not, then what does? It is obviously impossible to give a properly detailed account of Kant's ethics here, let alone the evidence necessary to back it up, but it is important that I should at least outline my own view, however briefly.

According to Kant, values are of two kinds: "dignity" and "price": "What has a price can be replaced by something else as its *equivalent*; what on the other hand is raised above all price and therefore admits of no equivalent has a dignity." (*Ak*. 4, 434) Dignity is, Kant says, "unconditional and incomparable" (*Ak*. 4, 436) in contrast to "price" in which trade-offs can be made. Only one thing, however, has "dignity": "morality, and humanity insofar as it is capable of morality, is that which alone has dignity." (*Ak*. 4, 435) Another term that Kant uses for whatever it is that has dignity is "*Persönlichkeit*". In the *Metaphysics of Morals*, at *Ak*. 6, 462, after "dignity", Kant

writes "*Persönlichkeit*" in parentheses. "*Persönlichkeit*" should undoubtedly be translated as "personhood", although the Cambridge edition translates it as "personality". "Personhood" is an aspect of human beings that transcends the empirical realm ("personhood, that is, freedom and independence from the mechanism of the whole of nature" (*Ak.* 5, 87)) and makes us, as it were, citizens of two worlds ("so that a person as belonging to the sensible world is subject to his own personhood insofar as he belongs to the intelligible world" (*Ak.* 5, 87)). It is from the inner, intrinsic value of "personhood" or "humanity in our person" that all other values descend.

Yes, you might say, but how? If "personhood" is a transcendental inner kernel that all of us carry within us, then it is, it seems, something inert and immune to human action. It can't be increased, diminished or destroyed. How, then, does it act as a foundation for subsidiary values like choice and happiness? How is it supposed to guide our actions? The immediate answer is, I think, clear. Dignity, personhood or whatever you like to call it, is not something that we can increase or pursue, not even something whose empirical existence we must defend. But it *is* something that we have an absolute duty to respect. Yet that, of course, might seem to do no more than kick the can down the road in front of us. We know how to respect things that can be violated, like the right to free speech, but how do we respect an indestructible inner transcendental kernel?

A proper treatment of this question would take us through the full corpus of Kant's ethical writings, but, in outline, my answer is that there are several apparently

different ways in which we must respect humanity in our (that is, of course, our own and others') persons. We must respect others' happiness and we must respect their choices (these two things actually come close together, since Kant's definition of happiness consists in things going according to our desires). In so doing we must respect human beings' equality; if what qualifies us for membership in the "kingdom of ends" is our possession of humanity in our persons, this is not something that any individual has in greater degree than any other. We must also act in ways that further the natural purposes that we find within ourselves (hence we must not let our capacities atrophy and devote our lives to "idleness, amusement, procreation – in a word, enjoyment" (*Ak*. 4, 423)). Finally, we must act in ways that are expressive of our respect for humanity – we must not allow ourselves to behave in a supine or submissive manner and we must not demean or disparage others. Between them, I think, these different ways of respecting humanity in our persons cover Kant's views about the different duties that we have.

But, beyond that, the idea that what is of intrinsic value in us is our moral *agency* points to something else fundamental. It is conventional in modern philosophy to distinguish between teleological and deontological ethics. Teleological ethics is said to be governed by a goal (or goals) to be attained or maximized, while deontological ethics acknowledges that there are constraints on such goals – values that may not be sacrificed in their pursuit. To the extent that "humanity in our person" embodies an inner, intrinsic value, Kant's ethics is plainly deontological in that sense. It is also deontological in a second, stronger sense, however: moral action itself has intrinsic value. That is not to say that there is nothing towards which morality is directed – as

stated above, we are required to act in ways that respect humanity in our persons. But moral agency is not merely a means to the realization or protection or even respecting of values; in moral action we exercise true freedom and so bring the highest kind of value into being. So Kantian ethics is action not outcome-oriented.

This, then, in briefest schematic outline, is a picture of Kant's moral theory to contrast with Parfit's, one which at least allows us to see Kant as someone who honours his own prescription that it is the highest duty of the philosopher to be "consistent" ("*consequent*") (*Ak*. 5, 24).

Yet, if this is really Kant, I imagine that Parfit will find it highly disappointing. If morality is a matter of respecting an inner, intrinsic value, how does that yield a criterion to tell us right from wrong? Can Kant give us no algorithm to compete with the Greatest Happiness Principle of the utilitarians? And if his ethics really is about respecting some inalienable inner transcendental kernel of our being and acting in ways that have value because they transcend the empirical world then he hardly looks like a pioneer of Non-Religious Ethics. Wouldn't an inconsistent secular thinker be better than a consistent one with those views?

In any case, how important is all of this? It is, of course, regrettable if Parfit's long interpretation of Kant fails to convince at least this reader, but how serious is that for his argument? He is, after all, an industrial-strength cognitivist realist about morality. So, you might think, what really matters in *On What Matters* is that claim about moral

truth and the historical picture of how awareness of it reveals itself once the miasma of theological illusion is dispersed is a mere framing narrative. To respond to these serious questions we must remind ourselves not only how much weight Parfit gives to the idea of coincidence in moral judgement, but say something more about the reasons why he does so.

V

As mentioned earlier, Parfit, like Sidgwick, is an intuitionist about moral judgements; that is, he believes that "We have *intuitive* abilities to respond to reasons and to recognize some normative truths" (*OWM,* 2, 544). Examples of such normative truths are that slavery is wrong and that we have reasons to prevent or relieve suffering (*OWM*, 1, 367). Such truths, Parfit claims, are necessary – true in all possible worlds (*OWM*, 2, 489).

That, however, is not the end of the story:

> We must also assess the strength of various conflicting reasons, and the plausibility of various principles and arguments, trying to reach what Rawls calls *reflective equilibrium*. This kind of intuitively-based reflective thinking is not only, as Scanlon writes, "the best way of making up one's mind about moral matters … it is the only defensible method." (*OWM*, 2, 544)

How do intuition and reflective equilibrium go together? Parfit does not explore the issue but it is important and underlies (I think) my deepest reservations about his project.

To explain Rawls's idea of "reflective equilibrium" it is easiest to start by comparing moral theory with a simple (Quine and Duhem would say, over-simplified) picture of empirical theory. On this picture, in the course of our experience of the empirical world we come up with particular observations and we develop generalizations – covering laws – to predict them. Such generalizations have logical priority over observations inasmuch as we can infer particular truths from them. Thus, if all swans are white, and Bert is a swan then Bert is white. On the other hand, if Ernie is a swan and he turns out to be black then it is the generalization "all swans are white" that must be given up. So while, on this simple picture, generalizations have logical priority over observations, epistemically, observations come first.

Consider now moral theory. Here too we have something like observations (first-order judgements) and something like general laws (moral principles) and they stand in similar logical relations. But what should we do if the two come into conflict? Imagine, for example, that we believe in the principle of free speech but also believe that an insulting, racist leaflet should be banned. In this case, so the idea of reflective equilibrium tells us, it is an open question which we give up – whether we revise our commitment to unrestricted free speech or decide that we must, however reluctantly,

allow the publication of the racist leaflet. What we must do is try to articulate our judgements and our principles (this is the "reflection" part) and bring them to a state of maximal systematic coherence and stability ("equilibrium").

From this brief presentation, it might sound as though reflective equilibrium is committed to anti-realism about morality. Surely, the idea that we move within our system of moral judgements, going up or down between principles and first-order judgements as seems best to us at the time, depends on the idea that there is nothing beyond that system to make to make such decisions right or wrong – nothing that, as it were, holds it fixed. But this is not quite correct. Certainly, reflective equilibrium is congenial to the anti-realist – it gives an explanation of how, even if there is nothing outside us towards which moral judgements are reaching out, the activity of moral theorizing can still be rational and illuminating. Yet reflective equilibrium does not *require* anti-realism – it can properly be agnostic about whether moral realism is true or not. Still, if one practises reflective equilibrium *and* is also a moral realist, then, depending on the nature of one's moral realism, the back-and-forth between principles and first-order judgements will be constrained.

Thus, if one is the kind of moral realist who sees human beings as having the intuitive capacity to make correct first-order moral judgements, reflective equilibrium will be a matter of developing moral theories to integrate those judgements into systematic structures. In other words, objective moral judgements will function very much like observations of the world do in the formation of empirical theories: they will act as anchor points and take priority. This is not the only way one could practise reflective

equilibrium as a moral realist, however. A moral realist might, for example, believe that objective moral truths were like mathematical axioms – fundamental starting points from which we might descend to particular cases. In that case the situation would be just the reverse and principles would have priority over first-order judgements.

VI

A point about the relationship between principles and first-order judgements in moral theory is extremely important here. Empirical generalizations allow us to infer particular facts – one predicts that a holiday in the Lake District is likely to have some rainy days because it rains a great deal in the Lake District, and so on. But in moral theory we need something more – we don't just want generalizations that cover our first-order judgements, we want principles that will *justify* those judgements. And this is built into the practice of morality itself. Whenever we make the decision to treat two (apparently) similar cases differently we need a reason to do so, not just a prediction that we will. We must explain why it is that the difference in question is a morally relevant one – that, in selecting someone for a job, for example, it is permissible to take into account the fact that she speaks French, but not the fact that she is female.

So what we might call a “covering law” model of moral principles won’t do at all. Imagine that we had discussed all the various “trolley problem” cases in the literature – the runaway trolleys that can be steered, the fat men standing on bridges, and all the rest – and had come to agreed and stable intuitions about how each case should be decided. Now imagine that we were to develop a rule – an extremely complicated one, no doubt – under which all of these cases would fall. Could we then sit back and consider our work done? I don’t believe so. We need to understand *why* it is that our judgements go in the direction they do in each case. In saying this, I don’t take myself to be asserting anything that puts me at odds with what those who participate in these debates are in fact doing. On the contrary, the distinctions between “doing” and “allowing”, between “wronging” and “harming”, the “doctrine of double effect” – even the “doctrine of triple effect” – are all, it seems to me, offered as ways of putting our fingers on subtle differences that are (or so it is claimed) morally significant.

The great moral theories of Kantianism and utilitarianism give clear (and very different) answers to the “why”-questions. As we have seen, for Kant, it is moral agency – “humanity in our person” – that sits at the core of his theory. It gives moral action its inherent value and is at the same time the “target” of such moral action – we must act in ways that respect humanity in our persons. Utilitarianism, in its classical form, offers a simple fundamental view: happiness is good, suffering bad. As far as Kant goes, of course, Parfit and I have radically different interpretations. If we start the reading of Kant through the eyes of the contractualist Rawls and his students and then revise his theory heavily in directions that are thought to make it more plausible

then perhaps he will emerge as more or less a modern contractualist moral theorist (although a very inconsistent one). But what about utilitarianism?

Here things are puzzling. Parfit claims that Kantianism implies "rule consequentialism", not utilitarianism. While utilitarianism has a very clear answer to the question of what makes an action right or wrong – that it advances or retards happiness – "consequentialism" leaves open what is to be advanced or retarded. To that extent, consequentialism is not so much a single moral theory as a family of moral theories (thus utilitarianism has been succinctly defined as "welfarist consequentialism"). Consequentialism's incompleteness, however, means that it lacks an account of what is, ultimately, of value. While Parfit makes many statements in the course of *On What Matters* that show that he is sympathetic to hedonism (most obviously, his repeated assertions that suffering is always bad and that no one could deserve to suffer) his official statement of the "Triple Theory" is that "Everyone ought to follow the principles whose being universal laws would make things go best" (*OWM*, 1, 25). How things would "go best" is left unspecified.

Yet I doubt that, on any consequentialist account of final value, "rule consequentialism" can be purely *consequentialist*. The problem is very simple. Take any given act that is prohibited by "rule consequentialism". Does it advance [whatever it is that is supposed to be the final end] or not when we include into our calculation the effect that breaking the rule under which it falls has on the strength and continuing operation of the rule itself? If it does not, then the act is prohibited on the grounds of simple consequences and "rule consequentialism' is just

consequentialism *tout court*. If, on the other hand, when all the consequences are taken into account, the act under review would advance [whatever it is that is supposed to be the final end] and yet that act is *still* held to be prohibited then there must be a reason other than [whatever it is that is supposed to be the final end] that justifies prohibiting it. So the allegedly consequentialist theory contains at least one non-consequentialist element, whatever final end one puts into that theory.

I have a similar difficulty with Parfit's response to questions of distribution. One of the oldest and most obvious objections to utilitarianism is that, since happiness is all that really matters, how happiness is distributed is, at best, a secondary issue. If we share psychologies and there is diminishing marginal utility then, yes, equality will be an advantageous policy, since it tends to advance happiness. Yet should this be a purely empirical matter? What if there are people who are particularly bad at converting resources into happiness? Should they lose out? Moreover, it is easy enough to come up with apparent examples in which a very large sacrifice by a single person (or a few) is outweighed by a small gain to a very large number.

Parfit addresses the issue of distribution in the second volume of *On What Matters*. We should, he says, incorporate into our moral theory what he calls the "Telic Priority View". "On the Telic Priority View, people's burdens matter more, doing more to make the outcome worse, the worse off these people are." (*OWM*, 2, 249) But how are we to understand this? Consequentialism tells us that outcomes are to be ranked by how far they produce whatever it is that the specific form of the theory counts as good or bad, whether it is pleasure or some other final good. But now the distribution of "burdens" is said to matter as well. It seems as if the consequentialist final good has

not just been left unspecified but also been given a curious elasticity – how much of it there is depends on where it is located – or something like a recursive quality – that, beyond first-order, immediate goodness, there is also a kind of goodness of goodness that depends on how that goodness is distributed. For if goodness is all that matters, how can the *location* of goodness matter – unless the location of goodness is itself a kind of good?

Putting together these three points – that, unless further specified, consequentialism does not give us a clear account of what the final end is to be pursued; that "rule consequentialism" adds a constraint upon the pursuit of whatever it is that is the final end that cannot be derived from the principle of consequentialism itself; and that the introduction of a distributive principle into consequentialism seems wholly *ad hoc* – then it looks to me as though, in this version at least, one of the three great moral theories that are held to converge lacks a clearly-articulated, coherent account of what is normatively fundamental.

Surely it makes a huge difference *why* people think that certain acts are wrong. People may agree that slavery is wrong, but does it not matter whether they do so, as Kantians, because slavery fails to respect humanity in our persons; as Lockeans, because it violates rights of self-ownership; or as consequentialists, because it leads to a net loss of happiness (or, indeed, because it deprives the slave of the necessary condition for accountability before a divine judge – which seems to have been a central consideration for those who actually abolished slavery)? I cannot imagine a satisfactory moral theory that does not give a persuasive answer to both the "what?"

and the “why?” of moral judgement. I want to know *why* advancing well-being is all that matters (if it does), why obeying rules should be right, even if that does not advance well-being, and I want to be given a reason why distribution matters – not just a highly abstract general principle that entails that it does.

Yet the following passage strongly suggests to me that Parfit himself sees things very differently:

> Some other moral disagreements are not about *which* acts are wrong but about *why* these acts are wrong, or what *makes* them wrong. Different answers are given by different systematic theories, such as those developed by Kantians, Contractualists and Consequentialists. Such disagreements do not directly challenge the view that we are able to recognize some moral truths. In defending this view, it is enough to defend the claim that, in ideal conditions, there would be sufficient agreement about which acts are wrong. Though we also have intuitive beliefs about why many acts are wrong, and about the plausibility of many systematic theories, we would expect there to be more disagreement about these other questions. As I have also argued, however, when the most plausible systematic theories are developed further, as they need to be, these theories cease to conflict. If that is true, these theoretical wars would end. (*OWM*, 2, 544)

In the light of it, I now think that the reason why the difficulties which seem so big to me do not trouble Parfit is because we are separated by a less obvious but, in the end, more profound disagreement about the nature of moral theory.

Parfit's "intuitively-based" reflective equilibrium is "bottom-up". That is, for him, what take priority are shared moral judgements. Moral theory will (it seems) be counted as adequate so long as it gives an accurate summary of the contours of how those moral judgements go together. If such principles should turn out to be extremely complex and, apparently, unmotivated, so be it. The requirements for convergence in moral theory thus become less demanding: two theories will count as "the same" provided that they are extensionally equivalent, whatever their account of what is, ultimately, of value. So objections that Parfit's picture of Telic Priority Rule Consequentialism does not contain a convincing, justifying account of the evaluative foundations of moral judgements (let alone one that coincides with Kant's) are, it seems, beside the point. If I am right about this and Parfit does not think a persuasive answer to why-questions is an essential part of moral theory then the divide between us is fundamental and unbridgeable.

VII

So, in the end, what really matters in *In What Matters*, I believe, is not agreement or disagreement as such but what agreement is meant to show: that we are capable of recognizing moral truths by intuition. As Parfit rightly says, disagreements about *why* acts are wrong or about *what makes them* wrong "do not directly challenge the view

that we are able to recognize some moral truths." (*OWM*, 2, 544) The critic might note, however, that, if someone claims that moral agreement results from people exercising a direct vision-like capacity to grasp the truth of the judgement in question, then it is very hard to see anything that would "directly challenge" that claim. Still, if people with different views about what, ultimately, makes acts right and wrong agree that slavery is wrong because all of them are intuitively recognizing it as a moral truth then they must at least be wrong in thinking that they are making that judgement as a result of their moral principles (which is, surely, what most of us think). Perhaps that should count as an indirect challenge.

My own view, by contrast, is that, even where there is agreement in judgement, conflict of principle may still be deep and real. Take our old friend, the trolley problem. Faced with that dilemma in its classic form, most (but by no means all) philosophers would choose to steer the trolley. What does this show? That they have all come to recognize an intuitive moral truth? Yet even when we judge that it is right to steer the trolley, most of us, surely, feel the force of claims on both sides.

On the one hand, is the simple, urgent fact that, whatever value may be at stake, there is apparently far more of it on the one side than the other. Not to acknowledge the compelling force of such aggregative claims and steer the trolley can seem perverse and even, perhaps (in giving excessive weight to our own agency) self-indulgent. On the other, however, are considerations telling us that we must not kill and that life and death are not the kind of things to be subjected to quantifying aggregation. These are deontological claims that derive from ideas about the intrinsic value of individuals

and convictions that actions themselves have worth that is not just to be measured by their outcomes, actual or likely.

These deontological claims have unmistakable religious roots. In its Kantian form, the idea that we are bearers of an inner, intrinsic kernel of value that all human beings (but only human beings) share in common clearly echoes the Christian doctrine of the soul. And the idea that the individual's exercise of moral agency is of absolute importance, independent of the outcomes that it brings about, makes sense in the context of a view of the world as a realm of "probation" in which individuals, at the end of their lives, will be held accountable before a just, but unyielding judicial deity. (Parfit, who finds the view that suffering could ever be deserved repellent, would, no doubt, be appalled by Kant's statement at the end of the *Metaphysics of Morals* that "it is from the necessity of punishment that the inference to a future life is drawn." (Ak. 6, 490)) Perhaps, then, a properly "Non-Religious Ethics" should purge itself of deontological elements and accept a thoroughgoing hedonistic consequentialism.

But how secular is pure hedonistic consequentialism? Taking up the "point of view of the universe" with Sidgwick and steering one's action according to what will maximize aggregate happiness here and everywhere, now and forever, looks like installing ourselves on the Deity's empty throne. Furthermore, if we come to see human beings as no more than a conjoined sequence of experiences, positive and negative, then, yes, death may lose its fearfulness – the coming to an end of that sequence should be no more terrible than the non-existence that preceded our births. But what would that mean for the value of life? Why would anyone worry about bringing one sequence of experiences to an end if that makes more, longer and better,

sequences possible? We have little experience of avowed atheists in positions of political power, but the record of those ruthless aggregators Lenin, Stalin, Mao and Pol Pot (to say nothing of Napoleon and Frederick the Great before them) should surely cause anyone to worry. Perhaps we would do better to cling on to the wreckage of our religious heritage.

I should like to be able to end this discussion by offering a solution – a way of founding respect for individuals on something other than a creaky, absolutist metaphysics and a way of adjudicating between competing claims on a basis of principle that is more plausible than aggregative hedonism. But it would be dishonest to pretend that I can. Still, that is the situation that I think that we find ourselves in, and that sets the problems for a Non-Religious Ethics.

Michael Rosen
Department of Government
Harvard University
mrosen@gov.harvard.edu
7.xi.2013